GNPAnnot Project 2010

Gaëtan Droc[1], Valentin Guignon[1], Franc-Christophe Baurens[1], Vincent Jouffe[1], Claire Poiron[1], Juliette Lengellé[1], Olivier Garsmeur[1], Mathieu Rouard[2],Stéphanie Bocs[1]

[1] UMR DAP, CIRAD, Montpellier [2] CfL, Bioversity, Montpellier

Michael Alaux[3], Leatitia Brigitte[3], Delphine Steinbach Samson[3], Erik Kimmel[3], Cyril Pommier[3], Isabelle Luyten[3], Nancy Terrier[5], Philippe Leroy[4], Hadi Quesneville[3]

[3] URGI, INRA, Evry [4] UMR GDEC, INRA, Clermont [5] UMR SPO, INRA, Montpellier

Joelle Amselem[6], Baptiste Brault[6], Adeline Simon[6], Victoria Dominguez Del Angel[6], Claire Hoede[6], Sabine Fillinger[6], Michel Meyer[6], Thierry Rouxel[6], Marc-Henri Lebrun[6]

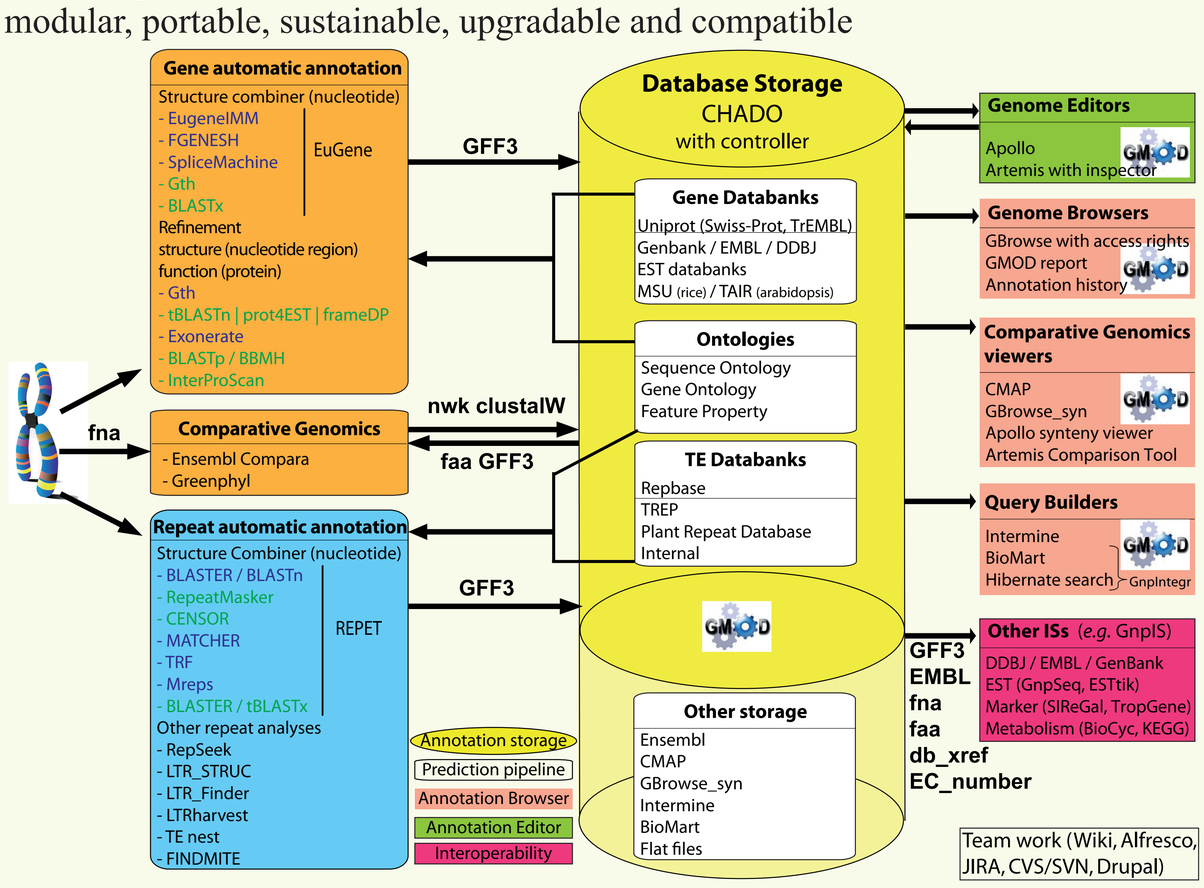
[6] UMR BIOGER, INRA, Versailles

Fabrice Legeai[7], Goulven Kerbellec[9], Olivier Collin[10], Jean-Pierre Gauthier[7], Emmanuelle d'Alençon[8], François Cousserans[8], Philippe Fournier[8], Denis Tagu[7]

[7] UMR BIO3P, INRA, Le Rheu [8] UMR BIVI, INRA, Montpellier [9] Korilog SARL, Muzillac [10] IRISA, INRIA, Rennes

Aims: GNPAnnot is a project on green genomics which intends to develop a community system of structural and functional annotation supported by comparative genomics and dedicated to plant and bio-aggressor genomes allowing both automatic predictions and manual curations of genomic objects.

Concept: The Community Annotation System (CAS) is user-friendly, generic, modular, portable, sustainable, upgradable and compatible



Results: Architecture of GNPAnnot CAS in three bioinformatics platforms

Component core	Montpellier	Versailles		Rennes	
Gene structure automatic annotation	EuGène	EuGène	TriAnnot		
Gene function & genome comparison	in-house pipeline	Funannot pipeline		MAUVE	
TE automatic annotation	REPET	REPET		REPET	
SGBD	Postgres Chado	Postgres Chado	MySQL BioDBSeqFeat	Postgres Chado	
Genome browser	GBrowse	GBrowse		GBrowse	
Genome editor	Artemis	Apollo		Apollo	
Synteny Viewer				Apollo	Cmap
Search & query builder		Biomart	Hibernate search	Apache Lucene	

GNPAnnot CAS resource statistics

Place	Subject	Unit	Organism	Genomic size (Mb)	Gene nb predicted	Gene nb curated	Gene nb current	TE nb predicted	TE nb curated	TE nb current
Montpellier	South & Tropical plants	DAP	Banana	7.13	1378	441	1298	3836	1279	2095
		CfL	Palm tree	0.27	43	30	41	5	5	9
			Sugarcane	1.30			133			113
Versailles	Wheat & grapevine	URGI SPO	Grapevine	480.00	26346			In progress		
		GDEC	Wheat 3B	18.21		175	175	10782	3222	3222
	Fungi	BIOGER	Botrytis	39.50	16360	1096		32	In progress	
			Leptosphaeria	44.90	12469	0		1850	472	
			Tuber	124.90	7496	1307		2520	0	
Rennes	Insects	BIO3P BIVI	Aphid	460.00	34821	1926	34547	498474	~800	498474
			Lepidopteran	4.00	1086	70	1086	2027	0	2027

GNPAnnot Portal

Home › Resources › Monocots Statistics

Musaceae Statistics — View | Edit | Revisions | Access control

GNPAnnot Musaceae Statistics

Unique Name	Clone Name	Alias	Accession Number	Length (bp)	Genes Count	Gene Density	%Gene	TE Count	TE Density	%TE	Gag-Pol Count	Project
scaffold_0001	Ma4001J14	MA4-1J14	AC226035	104637	18	4360	14%	8	13080	25%	6	45S rDNA locus
scaffold_0002	Ma4001C21	MA4-1C21		145629	15	8091	8%	18	8091	46%	3	26S rDNA locus
scaffold_0003	Ma4052E23	MA4-52E23		79629	15	4977	55%	9	8848	28%	1	RGA 08 locus
scaffold_0004	MaC091O16	MAC-91O16		92764	17	4882	38%	13	7136	39%	2	RGA 08 locus
scaffold_0005	Ma4071P20	MA4-71P20	AC226045	95242	13	7326	53%	1	95242	5%	0	Biotic stress
scaffold_0006	Ma4002G17	2G17	AC226036	37847	3	12616	21%	3	12616	37%	0	Cytogenetic marker
scaffold_0007	Ma4113F17	MA4-113F17		63169	9	5264	49%	0	n/a	n/a	3	RGA 08 locus
scaffold_0008	MbP008A02	MBP-8A2	AC226054	124336	24	5181	53%	0	n/a	n/a	0	Abiotic stress
scaffold_0009	MBP-36B13	MbP036B13		223694	38	5456	42%	20	11185	15%	3	RGA 08 locus

GBrowse

Change password... Logout sidbebocs

Showing 20 kbp from MbP036B13, positions 203,695 to 223,694

Instructions
Searching: Search using a sequence name, gene name, locus, or other landmark. The wildcard character * is allowed.
Navigation: Click one of the rulers to center on a location, or click and drag to select a region. Use the Scroll/Zoom buttons to change magnification and position.
Examples: scaffold_0001:1..20000, scaffold_0002:1..20000, scaffold_0003:1..20000, scaffold_0004:1..20000, Ma4002G17, MaC139M12.
[Bookmark this] [Upload your own data] [Hide banner] [Share these tracks] [Link to Image] [High-res Image] [Help]
Search
Landmark or Region: MbP036B13:203695..223694 Search
Data Source: Musa acuminata - GnpAnnot
Reports & Analysis: Annotate Restriction Sites Configure... Go
Scroll/Zoom: Show 20 kbp Flip
Overview
Region
Details
Eugene (cds)
Repet Manual
Exonerate TrEMBL(prot2genome)
BlastX Swiss-Prot
Name: MbP036B13_t380
Zoom in
Link
scaffold_0009:213305..218726
Detailed Report
Annotation History
Edit with apollo
Edit with artemis

History

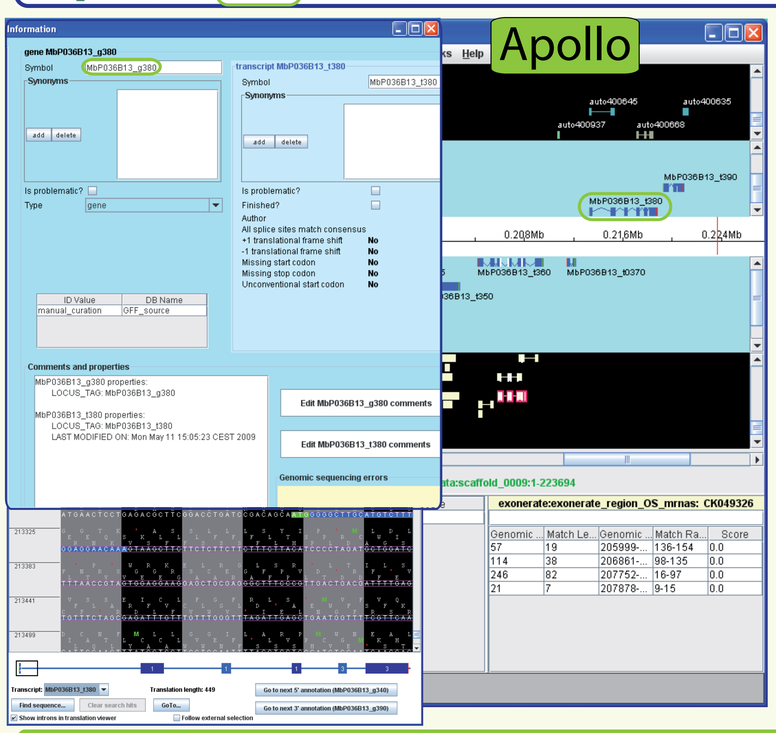
Date: 11:07 03/02/2010
Author: lengelle

Gene: 213305..218726
mRNA: 213305..218726
Exon: Join(213305..213335, 214994..215319, 216121..216256, 217097..217233, 217749..217872, 218131..218726)
/owner="gnpannot"
/note="MbP036B13_g380 Putative serine/threonine-protein kinase"
/annotator_comment="learning set; UTR to be annotated; homeologous gene MbP032N20cg440; as APK1B_ARATH, it belongs to RLCK family but is not orthologous to APK1B_ARATH http://greenphyl.cirad.fr/cgi-bin/family.cgi?id=36150 D:\Mes Donnees\projet\Musa\BAC\RGA\RGA08\phylogenie\greenphyl\MbP036B13g380.nwk TargetP does not predict transit peptide to Chloroplast as APK1B_ARATH D:\Mes Donnees\projet\Musa\BAC\RGA\RGA08\phylogenie\STK"
/inference="{EuGene rice 3.2; MSU/RGA:LOC_Os06g45280.1; RAP-DB:Os06g0663200; UniProtKB/TrEMBL:Q654A0; TAIR:At2g07180; UniProtKB/TrEMBL:Q9ZV76; TAIR:At2g28930; UniProtKB/Swiss-Prot:P46573; Uniprot/TrEMBL:C0P878_MAIZE; OS_mrnas:AK100479}"
/length="449 aa"
Product "Putative serine/threonine-protein kinase"
/completeness="complete"
/evidence_code="IC 2a"
/eC_number="2.7.11.1"
/locus_tag="MbP036B13_g380"

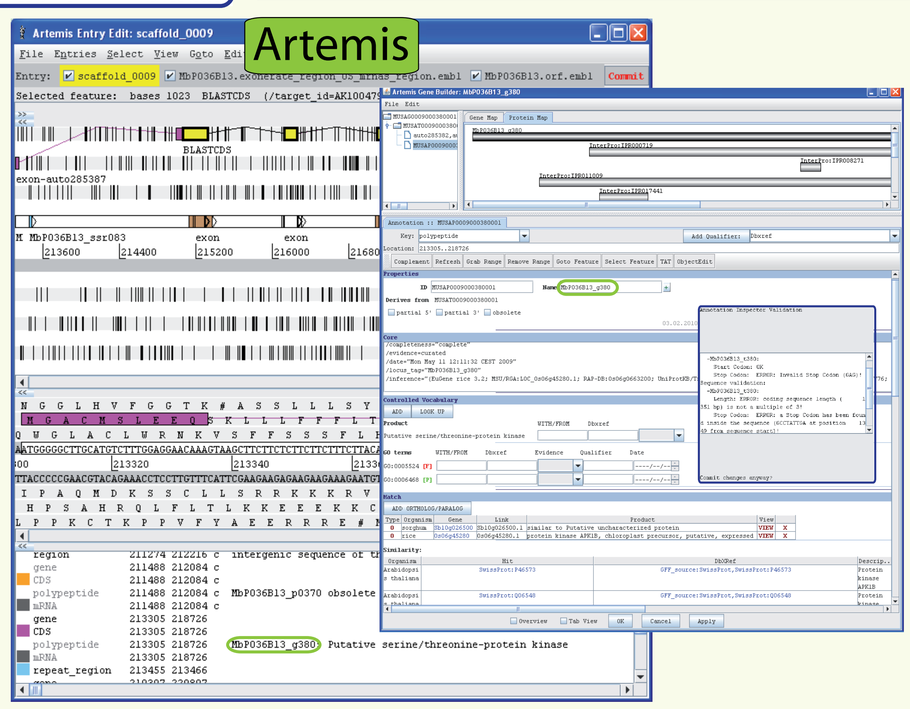
Date: 14:38 14/09/2009
Author: gnpannot

Gene: 213305..218726
mRNA: 213305..218726
Exon: Join(213305..213335, 214994..215319, 216121..216256, 217097..217233, 217749..217872, 218131..218726)
/owner="gnpannot"
/note="MbP036B13_g380 Putative serine/threonine-protein kinase"
/annotator_comment="homeologous gene MbP032N20cg440; as APK1B_ARATH, it belongs to RLCK family but is not orthologous to APK1B_ARATH http://greenphyl.cirad.fr/cgi-bin/family.cgi?id=36150 D:\Mes Donnees\projet\Musa\BAC\RGA\RGA08\phylogenie\greenphyl\MbP036B13g380.nwk TargetP does not predict transit peptide to Chloroplast as APK1B_ARATH D:\Mes Donnees\projet\Musa\BAC\RGA\RGA08\phylogenie\STK"
/inference="{EuGene rice 3.2; MSU/RGA:LOC_Os06g45280.1; RAP-DB:Os06g0663200; UniProtKB/TrEMBL:Q654A0; TAIR:At2g07180; UniProtKB/TrEMBL:Q9ZV76; TAIR:At2g28930; UniProtKB/Swiss-Prot:P46573}"
/length="449 aa"
Product "Putative serine/threonine-protein kinase"
/completeness="complete"
/evidence_code="3"

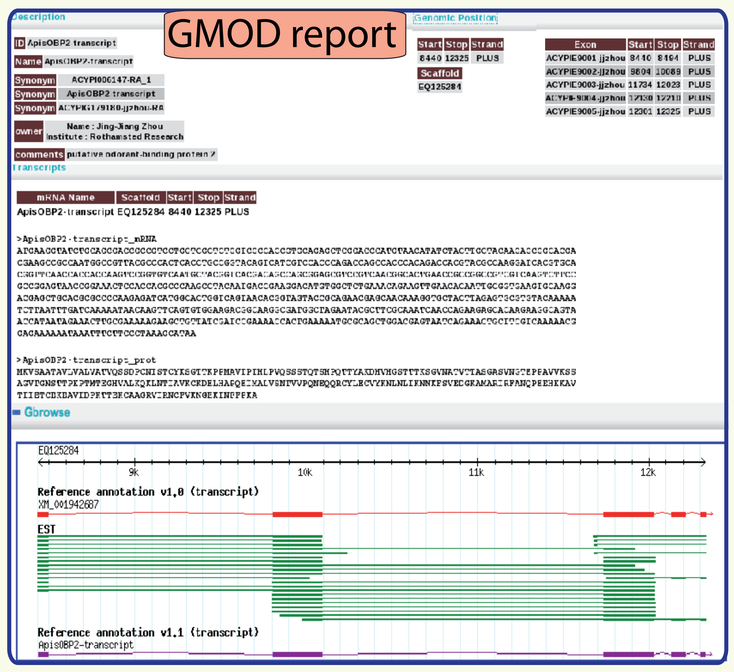
Apollo



Artemis



GMOD report



Done with plant, insect, fungal genomic sequences:

- Predictions of protein-coding genes and transposable elements
- CAS core roundtrips: Chado, GBrowse, Apollo, Artemis
- Feature, qualifier, value, annotation rule definitions
- Annotator training courses & manual curation of biological features
- GMOD report development
- Chado controller development to manage access rights, annotation inspector & history
- In collaboration with GnpIntegr project, advanced search user interface / query builders: Biomart, Hibernate search (lucene)
- Communications (posters, talks, Web site)

Ongoing work:

- JBrowse
- Annotation extractors, reconcilers & updaters (new genomic sequence, new gene annotation, other gene annotation set, new assembly of a genomic sequence)
- Comparative genomics
- Bioinformatics platform exchanges
- Integration of annotation history in the GMOD report
- Interoperability with other systems
- Communication (CECILL licences, publications)

References:

http://www.gnpannot.org/
http://www.gmod.org

Contact:

stephanie.sidibe-bocs@cirad.fr
michael.alaux@versailles.inra.fr
fabrice.legeai@rennes.inra.fr
joelle.amselem@versailles.inra.fr